PMC19

ESTIMATING SOCIODEMOGRAPHIC VARIABLES IN A PHARMACY DATASET: APPLYING DATA FROM US CENSUS 2000

LaFleur J, Ghate S, Ye X
University of Utah, Salt Lake City, UT, USA

OBJECTIVE: Pharmacy datasets are useful for evaluating drug costs, patient utilization, and patient adherence. Few have information on sociodemographic variables including race, education, income, and urban/rural designation of the patients' neighborhoods. We undertook to estimate those variables by linking USA census 2000 data with one such database. **METHODS:** We obtained census data tables for race, median income, education, and percentage of urban/rural residences by Zone Improvement Plan (ZIP) codes. Linking to the RxAmerica pharmacy claims database by 5-digit ZIP, we estimated sociodemographic variables for a cohort of adult new users of lipid-lowering therapy with ≥18-months continuous eligibility. For patients without a ZIP, pharmacy ZIP was used as a proxy. Four variables were generated to estimate race, median income, education level, and urban/rural designation of the patients' neighborhoods. **RESULTS:** A total of 29,667 patients met the inclusion criteria. Of these, 28,293 (95.4%) had a valid 5-digit ZIP that linked with one in the census tables; 97.0% of the ZIPs were derived from the patients' addresses. Among a sample of 19,458 patients, there were 4662 unique ZIPs; number of patients within each ZIP ranged from 1–125 (SD 11.9). The mean observation for income based on the median income variable reported in the census table was $45,924 (SD $15,965). The median observation for education was at least a high-school diploma; 25% of subjects had an observation corresponding to at least some college. Most of the patients had observations for race indicating they were from predominantly white neighborhoods (63.5% from neighborhoods that were ≥75% white). Most of the patients also had observations indicating they were from highly urbanized areas (74.0% from neighborhoods that were ≥75% urban). **CONCLUSION:** Census tables may be useful for estimating sociodemographic variables for pharmacy claims analyses. Future work should focus on the validity of these variables for estimating sociodemographic variables.

PMC20

METHODS TO SUMMARIZE COMPLICATED DATASETS CONTAINING STRUCTURED, NOMINAL DATA USING SAS

Zahedi H
University of Louisville, Louisville, KY, USA

Objective: The purpose of this study is to show the methodology for preprocessing and analyzing large health care databases. We consider working with large databases of clinical information such as National Inpatient Sample (NIS), and Thomson MedStat MarketScan data containing all patient claims in 40 million observations. **METHODS:** We can define a group of procedures and treat them as one episode to investigate the frequency of occurrence. In many studies, only the primary procedure and diagnosis are considered when there are more than one procedure and diagnosis columns, but important information could be in those other columns. In our database used for the study, there are fifteen procedure and fifteen diagnosis columns that we use to find episodes of patient care. We also combine information from multiple datasets: inpatient, outpatient, pharmacy information. Another approach is to consider a sequence of treatments on patients and to study the effectiveness of treatment by looking at this sequence for each patient. Studying the physician decisions and the results of them is interesting to many health care organizations. **RESULTS:** Powerful statistical software is required to work with large data files. We used SAS Enterprise Guide and the RXMATCH function to summarize codes defining a specific diagnosis, using multiple information sources. An alternative approach is to use SAS Text Miner. We combine columns using the CATX function. Then we use SAS Text Miner on the defined text string; the terms window in the output gives the frequency and number of documents. We use Text Miner features such as "Treating as equivalent terms", "Sorting" and "Filtering" to get summaries of different diagnosis or procedures. We successfully defined episodes of care. **CONCLUSIONS:** Preprocessing is an essential aspect of outcomes research. Dealing with multiple data sources is essential.

RESEARCH ON METHODS & CONCEPTUAL PAPERS—Modeling Studies

PMC21

COST-EFFECTIVENESS SENSITIVITY TO COST-EFFECTIVENESS CORRELATION: A SIMULATION STUDY

Muston D
Heron Evidence Development Ltd, Letchworth, Hertfordshire, UK

Incremental costs and effectiveness may be correlated if patients receiving the most health benefit from an intervention tend to be those who incur greater (or lesser) incremental cost. Such correlation has the potential to bias the synthesis of cost-effectiveness evidence, yet is infrequently reported. The objective of this study was to assess the sensitivity of an incremental cost-effectiveness ratio (ICER) for interventions A vs C, estimated indirectly from studies of A vs B and B vs C, to correlations between incremental cost and effect. Incremental costs and effects of interventions A vs B and B vs C were simulated from lognormal and Normal distributions respectively of given mean, standard deviation (SD) and cost-effectiveness correlation. The distribution of the ICER for interventions A vs C was calculated. Results from various parameter choices were compared. One parameter set choice was that the mean (SD) incremental cost of A vs B was $100,000 ($40,000) for a health benefit of ten(2) and that the incremental cost of B vs C was $200,000 ($50,000) for a health benefit of 12(2). For the parameter set described, the mean (SD) ICER of intervention A vs C assuming no correlation between incremental costs and health effects was $13,760 ($2637). Varying the two correlation parameters between −1 and +1 altered the mean ICER by up to $200 and the ICER's SD by up to $1679. The ICER's mean was fairly insensitive to the two correlation parameters but its SD was more sensitive. Correlations between cost and effectiveness can therefore be expected to be more important in questions of decision uncertainty rather than in the estimation with best existing knowledge of relative cost-effectiveness. More generally, sensitivity was greater when distributional variance and skewness of incremental cost and health benefit were greater.

PMC22

A MODEL TO EXAMINE THE EFFECT OF GUIDELINES ON OUTCOMES RESEARCH

Baser O[1], Wang L[2]
[1]STATinMED Research and University of Michigan, Ann Arbor, MI, USA, [2]STATinMED Research, Ann Arbor, MI, USA

OBJECTIVE: This paper introduces a method that combines the propensity score matching and interrupted (segmented) time series models to measure the effect of guidelines on outcomes measures. **MODEL:** Propensity score matching is used to balanced the groups before the trend is analyzed. The Kitchen sink approach is used for propensity score matching. Interrupted time series models applied over the matched sample. The time series